

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-056787

(43)Date of publication of application : 03.03.1995

(51)Int.Cl. G06F 12/00

(21)Application number : 06-141923

(71)Applicant : MICROSOFT CORP

(22)Date of filing : 23.06.1994

(72)Inventor : ZBIKOWSKI MARK
BERKOWITZ BRIAN T
FERGUSON ROBERT I

(30)Priority

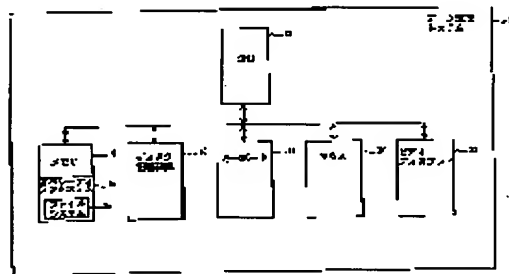
Priority number : 93 85483 Priority date : 30.06.1993 Priority country : US

(54) METADATA STRUCTURE AND METHOD FOR DEALING WITH THE SAME

(57)Abstract:

PURPOSE: To provide a file system for storing both data and metadata in a disk as the group of flows.

CONSTITUTION: A file system 26 stores data and metadata in a similar configuration. The lowest level of file data stored in a disk 16 is a flow constituting logically adjacent byte groups. The related flows seen in a file, directory, or sub-directory are stored in a node data structure in a variable size. The node data structure in the variable size is stored in the array of the fixed size packet of a disk space. The related node data structure is stored in a catalog data structure. The catalog data structure is stored in the array of the fixed size packet of a disk space. Next, the packet array of the fixed size packet of the disk space is stored as the flows.



LEGAL STATUS

[Date of request for examination] 20.06.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2000 Japan Patent Office

Best Available Copy

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-56787

(43) 公開日 平成7年(1995)3月3日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 12/00	5 2 0 J	8944-5B		
	E	8944-5B		

審査請求 未請求 請求項の数29 O L (全 12 頁)

(21) 出願番号 特願平6-141923

(22) 出願日 平成6年(1994)6月23日

(31) 優先権主張番号 08/085483

(32) 優先日 1993年6月30日

(33) 優先権主張国 米国 (U S)

(71) 出願人 391055933

マイクロソフト コーポレーション

MICROSOFT CORPORATI
ON

アメリカ合衆国 ワシントン州 98052-
6399 レッドモンド ワン マイクロソフ
ト ウェイ (番地なし)

(72) 発明者 マーク ズビコウスキ

アメリカ合衆国 ワシントン州 98072

ウッドインヴィル ノース イースト ワ
ンハンドレッドアンドセブンティエイス
プレイス 15817

(74) 代理人 弁理士 中村 稔 (外7名)

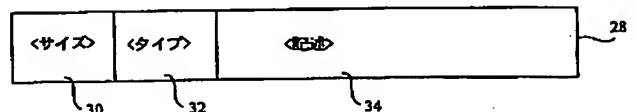
最終頁に続く

(54) 【発明の名称】 メタデータ構造体及びその取り扱い方法

(57) 【要約】

【目的】 データ及びメタデータの両方をディスクに流れのグループとして記憶するファイルシステムを提供する。

【構成】 ファイルシステムはデータ及びメタデータを同様の形態で記憶する。ディスクに記憶されたファイルデータの最低のレベルは、論理的に隣接するバイトグループを構成する流れである。ファイル、ディレクトリー又はサブディレクトリーに見られる関連する流れは、可変サイズのオノードデータ構造体に記憶される。可変サイズのオノードデータ構造体は、ディスクスペースの固定サイズバケットのアレーに記憶される。関連オノードデータ構造体はカタログデータ構造体内に記憶される。カタログデータ構造体は、ディスクスペースの固定サイズバケットのアレー内に記憶される。次いで、ディスクスペースの固定サイズバケットのバケットアレーが流れとして記憶される。



【特許請求の範囲】

【請求項1】 ディスク記憶装置と、オペレーティングシステムを実行する処理手段とを有するデータ処理システムであって、

(a) ディスク記憶装置において論理的に隣接するデータタイプを保持するための第1の可変サイズの流れデータ構造でデータを記憶し、

(b) ディスク記憶装置において論理的に隣接するデータバイトを保持するための第2の可変サイズの流れデータ構造でメタデータを記憶し、そして

(c) 上記流れデータ構造の各々に対し、その流れデータ構造がディスク記憶装置のディスクにいか記憶されるかを識別するタイプ識別子を含む流れ記述子を記憶する、という段階を備えたことを特徴とする方法。

【請求項2】 メタデータをディスク記憶装置において第2の可変サイズの流れデータ構造で記憶する上記段階は、ディスク記憶装置において関連データを保持する流れデータ構造に対するインデックスを第2の可変サイズの流れデータ構造で記憶する段階を更に備えた請求項1に記載の方法。

【請求項3】 ディスク記憶装置において関連データを保持する流れデータ構造に対するインデックスを第2の可変サイズのデータ構造で記憶する上記段階は、ディスク記憶装置において関連データを保持する流れデータ構造に対するBツリーインデックスを第2の可変サイズの流れデータ構造で記憶する段階を更に備えた請求項2に記載の方法。

【請求項4】 流れデータ構造の各々に対し流れ記述子を記憶する上記段階は、流れデータ構造の各々に対する流れ記述子を第1及び第2の各々の可変サイズの流れデータ構造で記憶する段階を更に備えた請求項1に記載の方法。

【請求項5】 ディスク記憶装置を有するデータ処理システムにおいて、

(a) ディスク記憶装置において論理的に隣接するデータバイトより成る流れデータ構造でデータの流れを記憶し、

(b) ディスク記憶装置において各流れデータ構造に対する流れ記述子を記憶し、各流れ記述子は、流れデータ構造がディスク記憶装置にいか記憶されるかを記述するタイプ識別子を含み、そして

(c) 関連データを記憶する流れデータ構造及びそれらに組み合わせられる流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶する、という段階を備えたことを特徴とする方法。

【請求項6】 ディスク記憶装置において流れデータ構造でデータの流れを記憶する上記段階は、ディスク記憶装置において流れデータ構造でメタデータの少なくとも1つの流れを記憶し、そしてディスク記憶装置において流れデータ構造でメタデータではないデータの少なくとも

も1つの流れを記憶するという段階を更に備えた請求項5に記載の方法。

【請求項7】 各流れデータ構造に対する流れ記述子をディスク記憶装置に記憶する上記段階は、各流れデータ構造に対する流れ記述子をディスク記憶装置において別の流れデータ構造で記憶するという段階を更に備えた請求項5に記載の方法。

【請求項8】 関連データを記憶する流れデータ構造及びそれらに組み合わせられた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、ファイルのデータを記憶する流れデータ構造及びそれらに組み合わせられる流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項5に記載の方法。

【請求項9】 関連データを記憶する流れデータ構造及びそれらに組み合わせられた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、ディレクトリーのデータを記憶する流れデータ構造及びそれらに組み合わせられた流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項5に記載の方法。

【請求項10】 関連データを記憶する流れデータ構造及びそれらに組み合わせられた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、サブディレクトリーのデータを記憶する流れデータ構造及びそれらに組み合わせられた流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項5に記載の方法。

【請求項11】 第1の可変サイズのデータ構造で記憶されず、関連データを記憶する流れデータ構造と、それに組み合わせられた流れ記述子とを、ディスク記憶装置において第2の可変サイズのデータ構造で記憶するという段階を更に備えた請求項5に記載の方法。

【請求項12】 第1の可変サイズのデータ構造及び第2の可変サイズのデータ構造をディスク記憶装置においてディスクスペースの固定サイズバケットのアレーに記憶するという段階を更に備えた請求項11に記載の方法。

【請求項13】 ディスク記憶装置におけるディスクスペースの固定サイズのバケットのアレーを流れデータ構造で記憶するという段階を更に備えた請求項12に記載の方法。

【請求項14】 ディスク記憶装置を有するデータ処理システムにおいて、

(a) データの流れをディスク記憶装置において流れデータ構造で記憶し、上記流れは、論理的に隣接するデータバイトより成り、

(b) 各流れデータ構造に対する流れ記述子をディスク記憶装置に記憶し、各流れ記述子は、流れデータ構造がディスク記憶装置にいか記憶されるかを記述するタイ

ブ識別子を含み、

(c) 関連データを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶し、そして

(d) 関連データを記憶する可変サイズのデータ構造のグループをカタログデータ構造に記憶する、という段階を備えたことを特徴とする方法。

【請求項15】 ディスク記憶装置において流れデータ構造でデータの流れを記憶する上記段階は、ディスク記憶装置において流れデータ構造でメタデータの少なくとも1つの流れを記憶し、そしてディスク記憶装置において流れデータ構造でメタデータではないデータの少なくとも1つの流れを記憶するという段階を更に備えた請求項14に記載の方法。

【請求項16】 各流れデータ構造に対する流れ記述子をディスク記憶装置に記憶する上記段階は、各流れデータ構造に対する流れ記述子をディスク記憶装置において別の流れデータ構造で記憶するという段階を更に備えた請求項14に記載の方法。

【請求項17】 関連データを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、ファイルのデータを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項14に記載の方法。

【請求項18】 関連データを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、ディレクトリーのデータを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項14に記載の方法。

【請求項19】 関連データを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において可変サイズのデータ構造で記憶する上記段階は、サブディレクトリーのデータを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において第1の可変サイズのデータ構造で記憶するという段階を更に備えた請求項14に記載の方法。

【請求項20】 第1の可変サイズのデータ構造に記憶されず、関連データを記憶する流れデータ構造と、それに組み合わされた流れ記述子とを、ディスク記憶装置において第2の可変サイズのデータ構造で記憶するという段階を更に備えた請求項14に記載の方法。

【請求項21】 第1の可変サイズのデータ構造及び第2の可変サイズのデータ構造をディスク記憶装置においてディスクスペースの固定サイズバケットのアレーに記憶するという段階を更に備えた請求項20に記載の方

法。

【請求項22】 ディスク記憶装置におけるディスクスペースの固定サイズバケットのアレーを流れデータ構造で記憶するという段階を更に備えた請求項21に記載の方法。

【請求項23】 ディスクスペースの固定サイズバケットのアレーを記憶する流れデータ構造に対する流れ記述子を、関連データを記憶する可変サイズデータ構造の1つに記憶するという段階を更に備えた請求項22に記載の方法。

【請求項24】 ディスクスペースの固定サイズバケットのアレーの所定の固定サイズバケットにカタログデータ構造を記憶する段階を更に備えた請求項22に記載の方法。

【請求項25】 ディスク記憶装置を有するデータ処理システムにおいて、

(a) データの流れをディスク記憶装置において流れデータ構造で記憶し、上記流れは、論理的に隣接するデータバイトより成り、

(b) 各流れデータ構造に対する流れ記述子をディスク記憶装置に記憶し、

(c) 関連データを記憶する流れデータ構造及びそれらに組み合わされた流れ記述子をディスク記憶装置において各可変サイズのデータ構造で記憶し、各可変サイズのデータ構造は、それに組み合わされた識別子を有し、

(d) 組み合わされた識別子を有する可変サイズのデータ構造をディスク記憶装置においてディスクスペースの固定サイズバケットのアレーに記憶し、

(e) 上記アレーにおけるバケットに対するバケット識別子を指定するエントリーを保持するマッピング構造をディスク記憶装置に記憶し、上記エントリーは上記可変サイズのデータ構造の識別子によってインデックスされ、そして

(f) 上記マッピング構造を使用し、アレーにおける上記可変サイズのデータ構造の1つを、その識別子が与えられると、位置決めする、という段階を備えたことを特徴とする方法。

【請求項26】 上記アレーは、ディスク記憶装置において流れデータ構造の1つで記憶され、そして上記アレーを記憶する流れに対する流れ記述子がディスク記憶装置に記憶される請求項25に記載の方法。

【請求項27】 上記マッピング構造は、ディスク記憶装置において流れデータ構造の1つで記憶され、そして上記マッピング構造を記憶する流れに対する流れ記述子がディスク記憶装置に記憶される請求項25に記載の方法。

【請求項28】 上記アレーを記憶する流れに対する流れ記述子及び上記マッピング構造を記憶する流れに対する流れ記述子は、可変サイズのデータ構造の選択された1つで記憶される請求項27に記載の方法。

【請求項29】 上記選択された可変サイズのデータ構造は、上記アレーのバケットの1つに記憶される請求項27に記載の方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は一般にデータ処理システムに係り、より詳細には、ファイルシステムによりディスクにデータを記憶する方法に係る。

【0002】

【従来の技術】 従来のファイルシステムは、ファイルデータをディスクに効率的に記憶する点で問題がある。従来の多くのシステムは、全てのデータをディスクに単一サイズの記憶単位で記憶する手法を採用している。不都合なことに、この手法は、ディスクにファイルデータを効率良く記憶するものではない。特に、ファイルデータはサイズが変化し、従って、所定の記憶単位サイズに良好に一致するものではない。従来の他のシステムは、多数の異なるフォーマットの1つを採用するオプションをユーザに与えるものである。ファイルデータがユーザに使用できるようになる前に、どのフォーマットを採用するか判断しなければならない。その結果、ユーザによるフォーマットの選択は単なる推測となり、実際のファイルデータとうまく対応しないことがしばしばある。従って、ファイルデータが効率的に記憶されないことが頻繁である。

【0003】

【発明が解決しようとする課題】 従来のオペレーティングシステムでは、各ファイルに固定数のディスクスペースのブロックが割り当てられ、ファイルデータと、ファイルに関する制御情報を記憶する。これらのブロックは、単位を素早く割り当てたり割り当て解除したりできるように固定サイズにされている。ファイル及びファイルデータに対する制御情報は、しばしば可変サイズとされ、これが少なくとも2つの問題を課することになる。第1に、ファイルデータ及び／又は制御情報が1つのブロックを埋めるに充分なほど大きくないときには、割り当て単位内のディスクスペースが浪費される。第2に、ファイルデータ及び／又は制御情報が1つのブロックに記憶するには大き過ぎるときには、それが多数のブロックに記憶されねばならず、データがどこに記憶されるかを指定するために多数のポインタが維持されねばならない。このようなポインタを維持しそして使用することは、かなり厄介であることが多い。

【0004】 従来のファイルシステムでは、データとメタデータ（即ち、他のデータの記憶を記述するデータ）とが個別のエントリーとして処理されている。特に、データとメタデータは、従来のファイルシステムでは異なるフォーマットで記憶されている。更に、データ及びメタデータに対して作用する個別のツールが設けられている。データとメタデータをこのように分けて考える結

果、オーバーヘッド及び複雑さが増大している。

【0005】

【課題を解決するための手段】 本発明の第1の特徴によれば、ディスク記憶装置と、オペレーティングシステムを実行する処理手段とを有するデータ処理システムにおける方法が実施される。この方法において、データは、ディスク記憶装置に、第1の可変サイズの流れデータ構造で記憶される。メタデータは、ディスク記憶装置に、第2の可変サイズの流れデータ構造で記憶される。各々の流れデータ構造に対して流れ記述子が記憶される。この流れ記述子は、流れデータ構造がディスク記憶装置内のディスクにいかかに記憶されるかを識別するタイプ識別子を含んでいる。

【0006】 本発明の別の特徴によれば、データの流れは、ディスク記憶装置に流れデータ構造で記憶される。データの流れは、論理的に隣接するデータバイトから形成される。各流れデータ構造に対しディスク記憶装置に流れ記述子が記憶される。各流れ記述子は、流れデータ構造がディスク記憶装置にいかかに記憶されるかを記述するタイプ識別子を含んでいる。関連データを記憶する流れデータ構造は、それらに組み合わせられた流れ記述子と共に、第1の可変サイズのデータ構造に記憶される。

【0007】 本発明の更に別の特徴によれば、データの流れは、ディスク記憶装置に流れデータ構造で記憶される。各流れデータ構造ごとに流れ記述子がディスク記憶装置に記憶される。各流れ記述子は、流れデータ構造がディスク記憶装置にいかかに記憶されるかを記述するタイプ識別子を含む。関連データを記憶する流れデータ構造は、それに組み合わせられた流れ記述子と共に、ディスク記憶装置に可変サイズのデータ構造で記憶される。関連データを記憶する可変サイズのデータ構造のグループは、カタログデータ構造で記憶される。

【0008】 本発明の更に別の特徴によれば、ディスク記憶装置を有するデータ処理システムにおける方法が実施される。この方法では、データの流れはディスク記憶装置に流れデータ構造で記憶される。流れは、論理的に隣接するデータバイトによって形成される。各流れデータ構造ごとに流れ記述子がディスク記憶装置に記憶される。関連データを記憶する流れデータ構造は、それらに組み合わせられた流れ記述子と共に、ディスク記憶装置に各々可変サイズのデータ構造で記憶される。各可変サイズのデータ構造は、それに組み合わせられた識別子を有している。可変サイズのデータ構造は、ディスク記憶装置におけるディスクスペースの固定サイズの識別可能なバケットのアレーに記憶される。このアレー内のバケットに対するバケット識別子を指定するエントリーを保持するマッピング構造がディスク記憶装置に記憶される。上記エントリーは、可変サイズデータ構造の識別子によってインデックスされる。このマッピング構造は、その識別子が与えられると、上記アレー内の可変サイズデータ

構造の1つを位置決めするのに使用される。

【0009】

【実施例】本発明の好ましい実施例は、データとメタデータの両方をディスクに「流れ」のグループとして記憶するファイルシステムを提供する。「流れ」とは、データのバイトの論理的に隣接しランダムにアクセスできる可変サイズのアレーであって、ディスク上の論理記憶単位として働くものをいう。ファイルのデータへのほとんどのプログラムアクセスは、これらの流れによって行われる。各流れは、多数の異なる表示(representation)の1つにおいてディスクに記憶される。異なる表示の各々は、流れの特定のサイズ及び使用に良く適したものである。従って、各流れは、そのサイズに最も良く適した表示でディスクに記憶される。

【0010】各流れには流れ記述子が組み合わされ、これはファイルシステムに制御構造において記憶される。流れ記述子は、流れをアクセスしそして流れに関する情報を得るのに使用される。流れ記述子は、流れのデータを記憶する表示の記述を与える。関連データを保持するデータの流れは、オノード(onode)として知られているデータ構造へとカプセル化される。オノードとは、ファイル、ディレクトリ又はサブディレクトリにほぼ類似した可変サイズの構造である。関連オノードのグループは、次いで、オブジェクト記憶カタログ又はオブジェクト記憶として知られているデータ構造に記憶される。従って、データ及びメタデータ(オブジェクト記憶カタログのような)は、同様の形態でハイアラーキに記述され、これについては以下で詳細に説明する。

【0011】図1は、本発明の好ましい実施例によるデータ処理システム10のブロック図である。図1のシステム10は単一プロセッサシステムであるが、本発明は分散型システムのようなマルチプロセッサシステムでも実施できることが当業者に明らかであろう。データ処理システム10は、中央処理ユニット(CPU)12と、メモリ14と、ディスク記憶装置16と、キーボード18と、マウス20と、ビデオディスプレイ22とを備えている。ディスク記憶装置16は、ハードディスク及び他の形式のディスク記憶装置を含む。キーボード18、マウス20及びビデオディスプレイ22は、従来型の入力/出力装置である。

【0012】メモリ14は、オペレーティングシステム24のコピーを保持し、これはシステムに記憶されたファイルを管理するためのファイルシステムマネージャ26を含んでいる。オペレーティングシステム24は、オブジェクト指向のオペレーティングシステムである。ここに述べる本発明の好ましい実施例は、オペレーティングシステム24の一部として実施される。本発明の好ましい実施例は、オペレーティングシステム24の一部として説明するが、当業者であれば、本発明は、オペレーティングシステムとは別の他の形式のコードでも実

施できることが明らかであろう。

【0013】上記したように、本発明の好ましい実施例においては、流れは、多数の異なる表示で得られる。流れについての異なる表示を理解するために、流れに対して与えられる流れ記述子のフォーマットを検討することが有用であろう。図2は、流れ記述子28のフォーマットを示す図である。流れ記述子28は、本発明の好ましい実施例で使用できる流れの異なる表示の各々を記述することができる。流れ記述子28は、サイズフィールド30、タイプフィールド32及び記述フィールド34の3つのフィールドを含んでいる。サイズフィールド30は、流れのサイズをバイトで指定する値を保持する。タイプフィールド32は、流れのタイプを指定し、そして記述フィールド34は、流れの記述(即ち、記述フィールド34の形式)を保持する。これらフィールド30、32及び34に保持された値は、それに関連する流れの表示と共に変化する(以下で詳細に述べる)。

【0014】「極小の流れ」は、本発明の好ましい実施例に使用できる流れの第1の表示である。この極小の流れは、記憶媒体(即ち、ディスク記憶装置16のディスク)の割り当て単位に対して非常にサイズの小さいデータを記憶するのに使用されるものである。記憶媒体の「割り当て単位」とは、ファイルを記憶するために割り当てられるディスク記憶装置16のディスクメモリスペースの基本単位を指す。例えば、FATベースのファイルシステムでは、最小割り当て単位はディスクセクタである。不都合なことに、このセクタは、システム10で形成される流れデータよりも相当に大きいことがしばしばある。図3は、極小の流れに対する流れ記述子28のフォーマットを示している。サイズフィールド30は流れのサイズを指定する値を保持し、そしてタイプフィールド32は、流れが極小の流れであることを指定する。記述フィールド34は、流れのデータを保持し、従って、流れのデータの即時表示を与える。この即時表示は、少量のデータを記憶するための非常に効率的な手段を形成する。特に、データは、流れ記述子に直接的に合体されて、容易に且つ速やかにアクセスできるようにされる。

【0015】本発明の好ましい実施例に使用できる別の表示は、小さな流れである。「小さな流れ」とは、データの単一のイクステントで記憶される流れである。イクステントとは、割り当て単位の変換サイズの隣接した延びである。流れのデータは、流れ記述子に直接記憶するには大き過ぎるのでイクステントに記憶される。この小さな流れに対する流れ記述子28のフォーマットが図4に示されている。タイプフィールド32は、それに関連した流れが小さな流れであることを指定し、そして記述フィールド34は、流れのデータが記憶されるイクステント42を記述するイクステント記述子36を保持する。イクステント42は、ディスク記憶装置16のディ

スクに記憶される。イクステント記述子36は、2つのサブフィールド38及び40を含む。サブフィールド38はイクステント42の長さを指定する値を保持し、そしてサブフィールド40は、イクステントのディスクアドレス（即ち、イクステントがディスクの論理アドレススペースのどこに位置しているか）を保持する。

【0016】本発明の好ましい実施例に使用できる第3の表示は、「大きな流れ」である。この大きな流れは、多数のイクステントに記憶される流れである。この大きな流れは、多量のデータを有する流れを記憶するのに適している。図5は、このような大きな流れに対する流れ記述子28のフォーマットを示している。タイプフィールド32は、流れが大きな流れであることを指定する。記述フィールド34は、イクステント記述子を保持する流れ44を記述する第2の流れ記述子43を保持する。この第2の流れ記述子43は極小の流れを記述し、そしてイクステント記述子の流れ44を保持する記述フィールド34'を含んでいる。イクステント記述子36'、36''及び36'''は、図4について述べたイクステント記述子36と同じフォーマットを有する。その結果、単一の流れ34'によって多数のイクステント42'、42''及び42'''が記述される。イクステント記述子の数があまりに多過ぎる場合には、第2の流れ記述子43は、極小の流れではなくて小さな流れを記述する。更に、イクステント記述子の数が小さな流れに対して多過ぎる場合には、第2の流れ記述子43は、大きな流れを記述する。大きな流れは、一般に、ディスクスペースの大きな隣接ブロックが得られない場合に使用される。この大きな流れは、多量のデータをディスクに対して分散されたイクステントにおいて単一の流れとして容易に記憶するものである。その結果、大きな流れは、ディスクが更に細分化されそして流れが成長するときにも、良好にスケールアップされる。

【0017】流れのサイズが成長するときには、流れの効率的な記憶を容易にするために、流れの表示が流れ表示のハイアラキを登るように促される。ハイアラキは、極小の流れ、小さな流れ及び大きな流れを含む。流れは、極小の流れから小さな流れへそして大きな流れへと促進される。一般に、上記したように、流れに含まれたデータの量及び細分の量に基づいて流れに最も適した表示が選択される。

【0018】本発明の好ましい実施例に使用できる4つの基本的な流れの形式を以上に説明した。流れ記述子28のタイプフィールド34は、流れ内に記憶されるデータの特殊な記述を指定するのに使用できる。図6は、流れが圧縮データを保持するときの流れ記述子28のフォーマットの一例である。タイプフィールド32は、流れのデータが圧縮されることを指定する値を保持し、一方、記述フィールド34は、圧縮データに対する流れ記述子を保持する。記述フィールド34に保持される流れ

記述子は、流れに含まれるデータの量に基づいて、極小の流れ、小さな流れ、又は大きな流れである。

【0019】図7は、流れ記述子が暗号データの流れを記述するときの流れ記述子28のフォーマットを示している。タイプフィールド32は、流れが暗号データを保持することを指定する値を保持する。記述フィールド34は、暗号データのための流れ記述子を保持する。又、流れ記述子は、暗号キーの値50も含む。この暗号キーの値50は、流れに記憶されたデータを暗号解読するのに使用できる。記述フィールド34に保持された暗号データに対する流れ記述子は、極小の流れ、小さな流れ、又は大きな流れである。

【0020】データの特殊な記述を指定するようにタイプフィールド32を使用する別の例が図8に示されている。図8は、小規模のトランザクションに対する流れ記述子28を示している。小規模のトランザクションとは、データベースのデータに対する変更が記録されるが、影響を受ける全てのデータの変更に関連したオーバーヘッドを被ることを保証するに十分な数の他の変更が生じるまでデータを直接的に変更しない場合を指す。タイプフィールド32は、記述フィールド34が小規模なトランザクションに対するデータを保持することを指定する。記述フィールド34は、第1の流れ記述子52と、第2の流れ記述子54を保持する。第1の流れ記述子52は、データの元の流れを記述する。第2の流れ記述子54は、元の流れに対して行われた変更を指定する流れを記述する。データの元の流れは、第2の流れに保持された変更を実施することによって更新される。

【0021】図9は、複製データを保持する流れに対する流れ記述子28の例を示す。データのロスが破壊的な結果を招かないように、データをしばしば複製しなければならない。特に、システムが多数のコピーをディスクに維持するところの選択されたデータ構造体がある。このような場合に、データ構造体は、ディスク上の2つの異なる位置にコピーされる。流れ記述子28は、第1の流れ記述子56と、第2の流れ記述子58をその記述フィールド34に含んでいる。第1の流れ記述子56は、第1のデータコピーを保持する第1の流れを記述し、そして第2の流れ記述子58は、別のデータコピーを保持する第2の流れを記述する。タイプフィールド32は、流れが複製データを含むことを指定する値を保持する。

【0022】本発明の好ましい実施例では、流れに関する情報が図10に示すようなフィールド記述子60に記憶される。このフィールド記述子60は、流れIDを保持する流れIDフィールド62を含んでいる。この流れIDは、オノード（以下で詳細に述べる）内の流れを独特に識別する4バイト長さの識別番号である。更に、フィールド記述子60は、フラグビットを保持するフラグフィールド64を備えている。フィールド記述子60の最終フィールドは、流れに対する流れ記述子28であ

る。

【0023】流れは、関連ファンクションに基づいて「オノード」にグループ分けされる。オノードはオブジェクトの論理表示に対応し、典型的に、ファイル、ディレクトリー又はサブディレクトリーを構成する全ての流れを保持する。各オノードは、これに含まれる流れの可変サイズの集合体を記述するに必要な情報を含む。

【0024】図11は、オノード66のフォーマットを示す図である。各オノード66は、関連ファンクションの流れを保持している。各オノード66は、次のフィールドを含む。即ち、長さフィールド68と、ワークIDフィールド70と、フラグフィールド72と、クラスIDフィールド73と、フィールド記述子60（図10に示すような）のアレーを保持するフィールド76とである。長さフィールド68は、オノードの長さを指定する値を保持し、一方、ワークIDフィールド70は、以下で詳細に述べるように、ワークIDマッピングアレー104（図14）へのインデックスを保持する。ワークIDは長さが4バイトである。フラグフィールド72（図11）はフラグビットを保持し、そしてクラスIDフィールド73は、オノードに対するクラスIDを保持する。フィールド76は、フィールド記述子60のバックアレーを保持し、これは、オノード66と共に保持される流れの各々に対するフィールド記述子60を含む。フィールド76のフィールド記述子のアレーに含まれる流れの数は、変化し得る。更に、フィールド76のフィールド記述子のアレーにおける各流れの長さも、変化し得る。従って、オノード66は可変サイズ構造である。オノード66の可変サイズ特性は、ディスク記憶装置16のディスクにおける割り当て単位での内部分割を最小にするよう助成する。

【0025】オノード66に関するあるデータは、オノードに直接組み込まれない。そうではなくて、このデータは、図12に示すように個別の流れに記憶される。流れ78は、関連オノード66（図11）に関する多数の異なる形式の状態情報を保持する。この状態情報は、フィールド86に保持されたタイムスタンプであって、オノード66が形成された時間を指定するタイムスタンプを含んでいる。フィールド88は、オノード66が変更された最後の時間を指定するタイムスタンプを保持する。同様に、フィールド90は、オノード66がアクセスされた最後の時間を指定するタイムスタンプを保持する。フィールド92は、オノード66のサイズを指定する値を保持し、そしてフィールド94は、オノードのオーナーに対する機密記述子を保持する。流れ78に保持された全ての情報は、関連オノード66に保持されたファイルデータを管理するのに有用である。

【0026】図12には、状態情報の第2の流れ80も示されている。この流れ80は、3つのフィールド96、98及び100を備えている。各オノード66は、

データ処理システム10のグローバルネームスペースに見ることができ、そしてこのグローバルネームスペースは、ルートオノード以外の各オノードが親ノードを有する論理ツリー構造である。フィールド96は、親オノードのワークIDを保持する。フィールド98は、オノード66のための全般的に独特のID（UID）を保持する。更に、フィールド100は、オノード66のクラスを指定するクラスIDを保持する。各オノード66には独特のクラスが組み合わせられる。特に、各オノード66は、特定のクラスのオブジェクトの例である。

【0027】状態情報の2つの付加的な流れ82及び84も記憶される。流れ82は、親に対するオノード66のネームを保持し、そして流れ84は、オノードのためのアクセス制御リスト（機密で使用する）を保持する。

【0028】流れ78、80、82及び84は、少なくとも次の2つの理由で別々に記憶される。第1に、この情報を別々に記憶すると、オノード66の平均サイズが減少される。流れ78、80、82及び84は、各オノードごとに記憶されない。その結果、オノードの平均サイズが減少する。第2に、この情報を別々に記憶すると、各オノード66からコードを検索するのに複雑なコードを必要とすることなく、情報の関連グループにプログラムアクセスすることができる。

【0029】オノード66（図11）は、オノードバケットアレー102（図13）に記憶される。このオノードバケットアレー102は、固定サイズバケットのアレーで構成された可変サイズデータ構造である。バケットのサイズ（例えば、4K）はデータ処理システム10のアーキテクチャに基づくものであるが、システム10のページサイズに一致してもよい。これらバケットは、図13の例では1ないしNと番号付けされている。アレー102の各バケットは、オノード66のバックされた組を含んでいる。図13の例では、バケット2は、単一のオノードのみを含み、一方、バケット1は、2つのオノードを含み、そしてバケット3は、3つのオノードを含むことが明らかである。

【0030】オノードバケットアレー102は、ファイルデータが移動、削除又は挿入されたときにシャフルされねばならないデータの量を最小にするために使用される。ディスクスペースのブロックを割り当て及び割り当て解除する粒度は固定とされる。換言すれば、メモリブロックは、固定サイズのバケットで割り当て及び割り当て解除される。そうではなくて、ファイルデータを記憶するために可変サイズの構造が使用された場合には、割り当ての粒度は固定されず、割り当ての粒度を非常に大きくすることができる。

【0031】効率化のために、オペレーティングシステム24（図1）は、関連オノード66（図11）をアレー102（図14）の同じバケットに記憶するか、又は

ディスク記憶装置16のディスク上で互いに接近したバケットに記憶する。この記憶戦略は、ディスク上でオノード66を探すシーク時間を最小にする。一般に、通常一緒にアクセスされるファイルは同じバケットに配置される。一緒にアクセスされるファイルは、例えば、共通のサブディレクトリー又はディレクトリーにあるファイルである。

【0032】オノードバケットアレー102(図13)は、流れとしてアクセスされる。システム10は、通常、多数のオノードバケットアレー102を記憶し、従って、オノードバケットアレーを保持する多数の流れを記憶する。内部では、オノード間の全ての参照は、ワークIDに基づいている。アレー102のバケット内でオノード66(図11)を位置決めするには、そのオノードに対するワークIDを探すことが必要である。ワークIDのマッピングアレー104(図14)は、ワークIDからオノードバケットアレー102のバケット番号へマップする。次いで、番号付けされたバケットがオノードに対し整合ワークIDでサーチされる。特に、特定ノード66のワークIDは、ワークIDマッピングアレー104へインデックスを与える。指定されたインデックスにおけるエントリーは、オノード66を保持するバケット番号を識別する。例えば、図14に示すように、ワークIDマッピングアレー104のエントリー106及び108は、関連するワークIDを有するオノード66がバケット1に保持されることを指定する値を保持する。

【0033】各オノード66(図11)は、フィールド記述子76のアレーに保持される流れの中に記憶されるネームインデックスの流れを含んでいる。ネームインデックスの流れは、オノード66内に含まれた流れ記述子に対するBツリーインデックスを保持する。ネームインデックスの流れは、流れの流れIDを、オノード66内に保持された流れ記述子を位置決めするためのキーとして使用する。Bツリーインデックスは、本発明と同日に出願された共通の譲受人に譲渡された「ファイルシステムにオブジェクトを効率的に記憶する方法(Efficient Storage of Objects in a File System)」と題する特許出願に詳細に示されている。この特許出願の開示を参考としてここに取り上げる。

【0034】流れの関連収集体がオノード66(図11)へと収集されると同様に、オノードの関連収集体がオブジェクト記憶カタログとして知られたデータ構造へ収集される。オノード66のワークIDは、オブジェクト記憶カタログ内のオノードを識別するためのベースとして働く。オブジェクト記憶カタログは、図15に示すようなオブジェクト記憶カタログオノード110によって記述される。オブジェクト記憶カタログオノード110は、長さフィールド68と、ワークIDフィールド70と、フラグフィールド72と、他のオノード66で

見つかったフィールド記述子76のアレーとを含んでいる。フィールド記述子76のアレーは、流れ記述子112、114及び116が記憶されたフィールド記述子を含む。

【0035】流れ記述子112は、ワークIDマッピングアレーの流れ104を記述する。流れ記述子114は、オノードバケットアレーの流れ102を記述し、そして流れ記述子116は、ネームインデックスの流れ117を記述する。ワークIDマッピングアレーの流れ104及びオノードバケットアレーの流れ102は、関連データを記憶するので、これらは共通のオノード(即ち、オブジェクト記憶カタログオノード110)に記憶される。このように、オブジェクト記憶カタログオノード110は、ワークIDマッピングアレーの流れ104、オノードバケットアレーの流れ102及びネームインデックスの流れ117を関係付けるビークルを形成する。オブジェクト記憶カタログオノード110は、ワークID 0におけるカタログを構成するオノードのグループ内に記憶される。ワークID 0とは、その定義により、オノードバケットアレーの流れ102の第1エレメントに常に配置される。ワークID 0における第1バケットは、既知の位置を有する区別されたバケットである。従って、バケット内に記憶されたデータ構造は容易に且つ簡単にアクセスされる。

【0036】本発明を好ましい実施例について説明したが、特許請求の範囲に規定された本発明の範囲から逸脱せずにその形態及び細部に種々の変更がなされ得ることが当業者に明らかであろう。

【図面の簡単な説明】

【図1】本発明の好ましい実施例によるデータ処理システムのブロック図である。

【図2】本発明の好ましい実施例に用いられる流れ記述子のフォーマットを示す図である。

【図3】本発明の好ましい実施例による極小の流れに対する流れ記述子の図である。

【図4】本発明の好ましい実施例による小さな流れに対する流れ記述子の図である。

【図5】本発明の好ましい実施例による大きな流れに対する流れ記述子の図である。

【図6】本発明の好ましい実施例による圧縮流に対する流れ記述子の図である。

【図7】本発明の好ましい実施例による暗号流に対する流れ記述子の図である。

【図8】本発明の好ましい実施例による小規模トランザクションに対する流れ記述子の図である。

【図9】本発明の好ましい実施例による複製データに対する流れ記述子の図である。

【図10】本発明の好ましい実施例によるフィールド記述子の図である。

【図11】本発明の好ましい実施例によるオノードの図

である。

【図12】図11のオノードについての適切な情報を保持するシステムの図である。

【図13】本発明の好ましい実施例によるバケットアレーの図である。

【図14】本発明の好ましい実施例によるワークIDマッピングアレー及びオノードバケットアレーの図である。

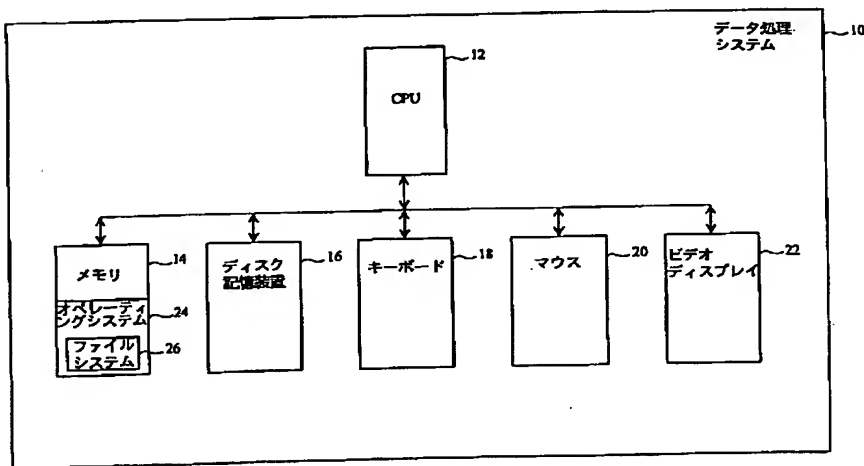
【図15】本発明の好ましい実施例によるオブジェクト記憶カタログの図である。

【符号の説明】

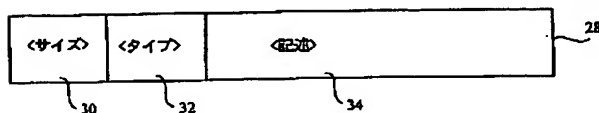
10 データ処理システム
12 中央処理ユニット
14 メモリ
16 ディスク記憶装置
18 キーボード
20 マウス
22 ビデオディスプレイ

24 オペレーティングシステム
26 ファイルシステム
28 流れ記述子
30 サイズフィールド
32 タイプフィールド
34 記述フィールド
38、40 サブフィールド
42 イクステント
43 第2の流れ記述子
60 フィールド記述子
66 オノード
68 長さフィールド
70 ワークIDフィールド
72 フラグフィールド
73 クラスIDフィールド
76 フィールド
78 流れ
80 第2の流れ

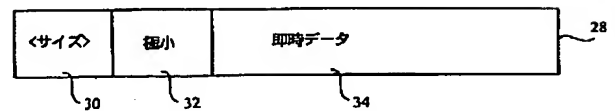
【図1】



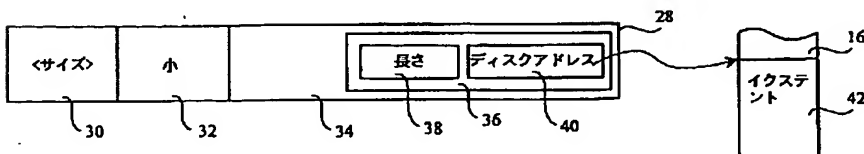
【図2】



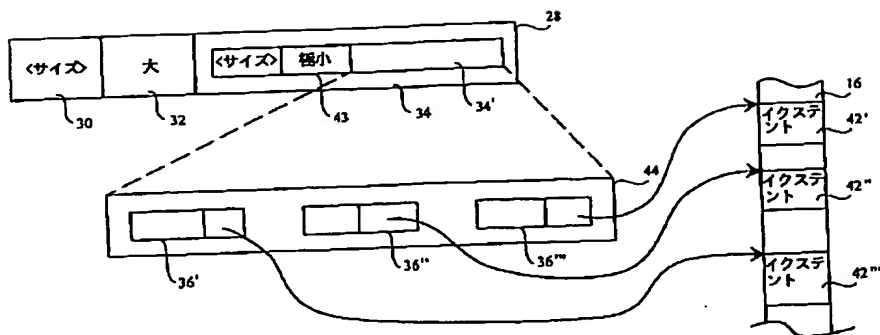
【図3】



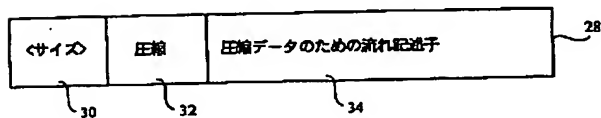
【図4】



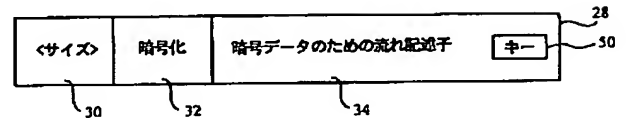
【図5】



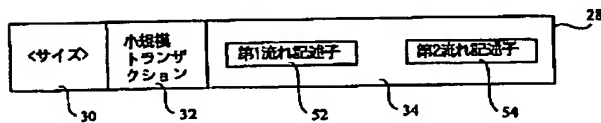
【図6】



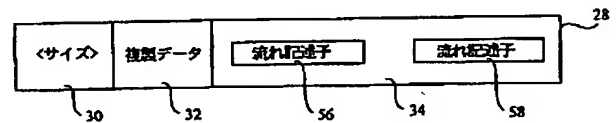
【図7】



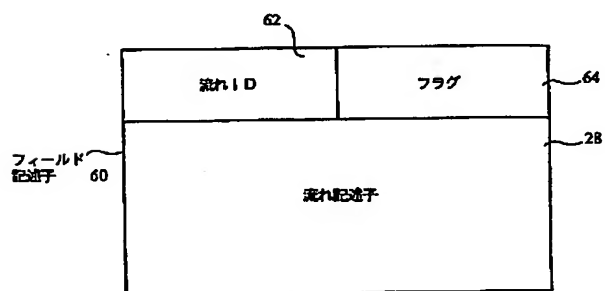
【図8】



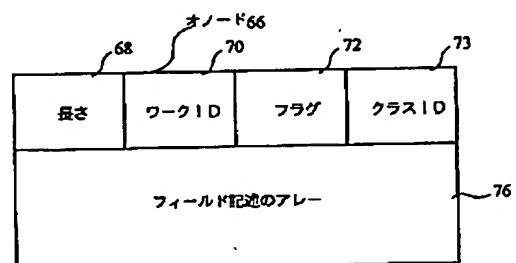
【図9】



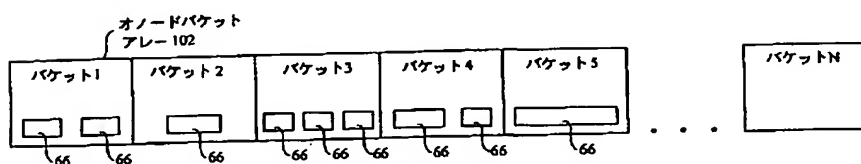
【図10】



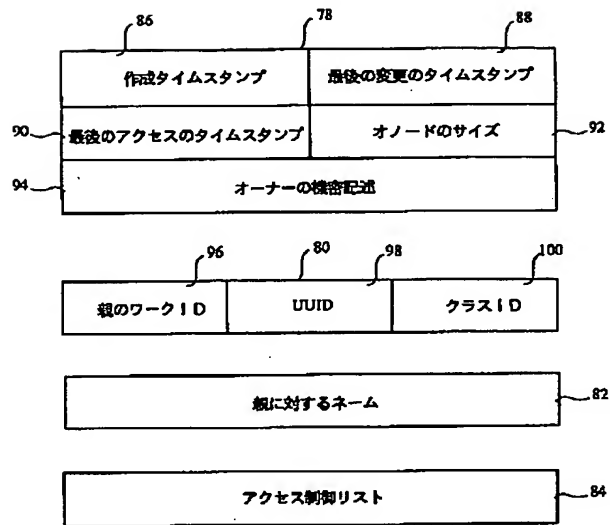
【図11】



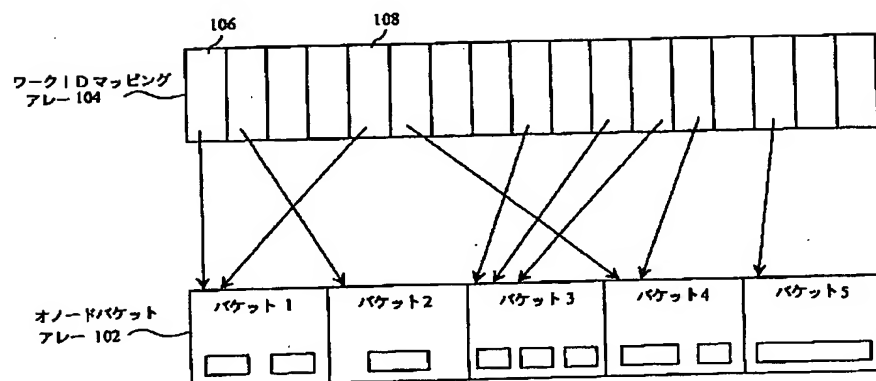
【図13】



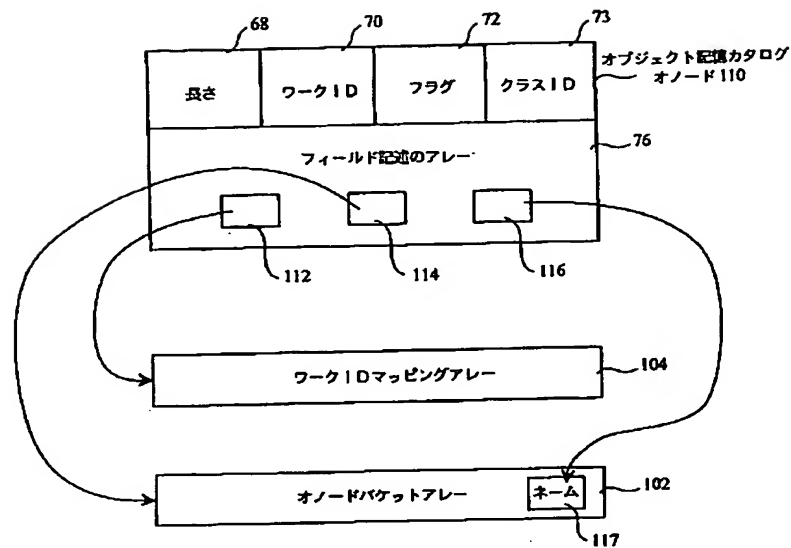
【図12】



【図14】



【図15】



フロントページの続き

(72)発明者 ブライアン ティー バークウィッツ
 アメリカ合衆国 ワシントン州 98007
 ベルヴィュー ワンハンドレッドアンドフ
 ォーティセカンド プレイス ノース イ
 ースト 3912

(72)発明者 ロバート アイ ファーガソン
 アメリカ合衆国 ワシントン州 98119
 シアトル ナインス アベニュー ウェス
 ト 2910

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.